AD-763 404

ON A UNIFIED THEORY OF ESTIMATION IN LINEAR MODELS

C. R. Rao

Purdue University

Prepared for:

Office of Naval Research

May 1973

1

PURDUE UNIVERSITY

DEPARTMENT OF STATISTICS

DIVISION OF MATHEMATICAL SCIENCES

41

On a Unified Theory of Estimation

in Linear Models *

by

C. R. Rao

Indian Statistical Institute

and

Indiana University

Department of Statistics
Division of Mathematical Science
Mimeograph Series #319

May 1973

Purdue University
W. Lafayette, Indiana
47907

*These research lecture notes were partly sponsored by the Office of Naval Research Contract N00014-67-A-0226-00014 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Preface

In March of 1973 Professor C. R. Rao gave a special series of colloquium lectures at Purdue University. The first three lectures were on a unified theory of estimation in the General Gauss Markov linear model. During the lectures, notes were taken of the material presented by Professor Rao. This report is a presentation of these lecture notes together with additional details of proofs which Professor Rao kindly supplied.

The responsibility for the correctness and accuracy of these notes lies with the note taking committee composed of Professors Cote and myself, Dr. T. Santner and Mr. A. K. Bhargava. The committee thanks Professor Rao for his patience in answering questions which arose during the preparation of these notes. Special thanks are also due to Mr. A. K. Bhargava for his substantial efforts in preparing these notes.

E. M. Klimko
Colloquium Chairman

1.1 INTRODUCTION

In a series of papers the lecturer developed two approaches towards a unified treatment of the General Gauss-Markoff (GGM) linear model $(Y, X\beta, \sigma^2 V)$ where V, the dispersion matrix of Y, may be singular and X may be deficient in rank. One is called the inverse partition (IPM) method which depends on the numerical evaluation of a g-inverse of a partitioned matrix. Another is an analogue of least square theory and is called unified least square (ULS) method.

It may be noted that Aitken's [1] approach (which is called generalized least squares) is applicable only when V is non-singular although the requirement that X is of full rank can be relaxed.

The aim of these lectures is to bring out the salient features of these two methods and to point out some interesting features of linear unbiased estimation when the dispersion matrix of the observations is singular.

1.2 STATEMENT OF THE PROBLEM

Consider the triplet

(1.2.1) $$(Y, X\beta, \sigma^2 V)$$

where Y is an nx1 vector of random variables, X is a given nxm matrix and β is an unknown mx1 vector. Furthermore,

$$E(Y) = X\beta$$

and $$D(Y) = \sigma^2 V$$

where σ^2 is unknown.

We refer to set up (1.2.1) as the General Gauss-Markoff (GGM) model. No assumption is made about R(V) and R(X) where R(·) denotes the rank of the matrix argument.

The problem is to estimate β and σ^2. An associated problem is that of testing hypotheses e.g. Test H_0: $P'\beta = w$, where P is a given kxm matrix, on the basis of the given model.

The classical method of solving the above problem is the method of least squares. Various types of difficulties can arise i.e. the parameters may not be independent and the variables may be related in the following sense:

(a) $R(X) < m$

(b) $R(V) < n$ ($|V| = 0$)

If neither of the above two difficulties is present then a solution to the problem of estimation of β is the $\hat{\beta}$ which minimizes (β is not estimable if $R(x) \neq m$)

$$Q = (Y-X\beta)'V^{-1}(Y-X\beta).$$

2. PRELIMINARIES

2.1 Notation

The following notation will be used throughout.

(a) The vector space generated by the columns of a matrix X is represented by $\mathcal{M}(X)$.

(b) The vector space orthogonal to $\mathcal{M}(A)$ is denoted by $\mathcal{M}(A^{\perp})$ where $A^{\perp}$ is a matrix of maximum rank with its columns orthogonal to the columns of A.

(c) If V is a n.n.d. (non-negative definite) matrix the expression

$$||p|| = (p'Vp)^{1/2}$$

where p is a vector is called the V-norm of p.

(d) The BLUE (best linear unbiased estimator) is the linear unbiased estimator with minimum variance.

(e) $(X:V)$ denotes a partitioned matrix and $R(X)$ the rank of matrix X. A matrix with all zero entries is denoted by 0.

2.2 Some Results on g-inverses of a Matrices

Def. 2.2.1 Let A be an mxn matrix. A g-inverse of A is an nxm matrix denoted by A^-, satisfying the condition

$$AA^-A = A$$

Generalized inverses have the following properties.

(a) $AA^-B = B \Leftrightarrow B = AK$ i.e. $\mathcal{M}(B) \subset \mathcal{M}(A)$.

Proof:

Sufficiency is obvious. To prove necessity choose

$$K = A^-B.$$

(b) Let A be of order mxn and let A^- be any g-inverse of A.

Then

(i) A general solution of the homogeneous equation

$$Ax = 0$$

is

$$x = (I-A^-A)z,$$

where z is an arbitrary vector;

(ii) a general solution to a consistent non homogeneous equation

$$Ax = y$$

is

$$x = A^-y + (I-A^-A)z,$$

where z is an arbitrary vector.

Proof:

(i) Note that this is equivalent to saying that the orthogonal space of $A' = \mathcal{M}(I-A^-A)$ which follows from the fact that

$$A(I-A^-A) = 0$$

and $$R(I-A^-A) = n-R(A).$$

(ii) follows since a general solution of $Ax = y$ is the sum of a particular solution of $Ax = y$ and a general solution of $Ax = 0$.

(c) The projection operator on $\mathcal{M}(x)$ is

(i) $P_X = X(X'X)^-X'$, which is unique for any choice of the g-inverse, when the inner product $(x,y) = x'y$, and

(ii) $P_X = X(X'\Lambda X)^-X'\Lambda$, which is unique for any choice of the g-inverse, then the inner product $(x,y) = x'\Lambda y$, Λ being a p.d. matrix.

Proof of (i): By property (a), we have $X'X(X'X)^-X' = X'$. Then

$$\therefore P_X = X(X'X)^-X'X(X'X)^-X' = X(X'X)^-X' = P_X$$

so that P_X is idempotent.

Further $[(X'X)^-]'$ is also a g-inverse of $X'X$. Then by uniqueness for choice of g-inverse

$$P_X' = X[(X'X)^-]'X' = X(X'X)^-X' = P_X$$

so P_X is symmetric. Thus P_X is the projection operator.

Proof of (ii): The proof is the same as in (1). We establish P_X is idempotent and ΛP_X is symmetric.

(d) $\mathcal{M}(X^\perp) = \mathcal{M}[I-(X')^-X']$.

Proof: Let $R(X') = r$.

Then $X'[I-(X')^-X'] = X'-X'(X')^-X' = X'-X' = 0$

Next we show that $R[I-(X')^- X'] = n-r$.

This follows easily from the fact that the matrices $[I-(X')^- X']$, I and $(X')^- X$ are all idempotent. Therefore

$$R[I-(X')^- X'] = \text{Trace}[I-(X')^- X'] = \text{Trace } I - \text{Trace } (X')^- X' = n-r$$

(e) Consider the equation

(2.2.1) $$AXA = A$$

Then four alternative representations of a general solution to (2.2.1) are, with P_A as the projection operator on $\mathcal{M}(A)$,

(i) $X = A^- + U - A^- AUAA^-$

(ii) $X = A^- + (I-A^- A)V+W(I-AA^-)$

(iii) $X = A^- + U - P_{A'} UP_A$

(iv) $X = A^- + W(I-P_A) + (I-P_{A'}) V$

where A^- is a particular g-inverse and U, V, W are arbitrary matrices.

Proof:

Verification of these identities is straightforward and left to the reader.

(f) The equation $AXB = C$ has a solution if and only if

(2.2.2) $$AA^- C B^- B = C.$$

In such a case a solution is given by

(2.2.3) $$X = A^- C B^- + Z-A^- A Z BB^-$$

where Z is arbitrary.

Proof:

Necessity of (2.2.2) follows from the fact that if the equations are consistent there exists a matrix X such that

$$AXB = C$$

Then $AA^- C B^- B = AA^- AXBB^- B = AXB = C$. Sufficiency is trivial since here $A^- CB^-$ is clearly a solution. Observe that X defined by (2.2.3) satisfies the equation

$$AXB = C.$$

Also any arbitrary solution X of this equation is obtainable through the formula (2.2.3) by a suitable choice of the matrix Z; for example,

$$Z = X - A^- C B^-$$

is such a choice. This shows that (2.2.3) provides the general solution.

(g) (A generalization of Fisher-Cochran's Theorem.)

Theorem 2.2.1. Let A_i be $m \times p_i$ matrices of rank r_i $i = 1,2,\ldots,k \ni \sum_{i=1}^{k} r_i = m$.

Then the following are equivalent:

(i) $A_i' A_j = 0 \quad i \neq j$

(ii) $I = \sum_{i=1}^{k} A_i (A_i' A_i)^- A_i'$

Proof: Rao and Mitra [4] prove a more general result, Theorem 2.8.1 on p. 33-34.

(h) Let V be a n.n.d. matrix and X be any given matrix. If there exists a matrix U such that

$$\mathcal{M}(V + XUX') = \mathcal{M}(V:X).$$

Then,

$$R[X'(V + XUX')^- X] = R[X'].$$

Proof:
See Lemma 5.2.2 for the proof.

(i) Def. 2.2.2. A matrix denoted by $A^-_{M(N)}$ is said to be a minimum N-norm inverse of A if

$$\hat{x} = A^-_{M(N)} y$$

is a solution of the consistent-equation

$$Ax = y$$

with the smallest N-norm (being defined as $\sqrt{x'Nx}$) where N is an n.n.d. matrix.

Remark 2.2.1

(i) $A^-_{M(N)}$ need not be unique

(ii) $\{A^-_{M(N)}\} \subset \{A^-\}$

(iii) $A^-_{M(N)} = G$

if and only if

$$AGA = A$$

$$(GA)'N = NGA$$

(j) Let $Ax = y$ be a not necessarily consistent equation then a matrix denoted by $A^-_{\ell(M)}$ is said to be M-least square inverse of A if

$$\hat{x} = A^-_{\ell(M)}\, y$$

minimizes the quadratic form

$$(Ax-y)'\, M(Ax-y)$$

where M is a p.d. matrix. $\hat{x}$ is called a M-least squares solution of $Ax=y$.

Remark 2.2.2

(i) $A^-_{\ell(M)}$ need not be unique

(ii) $\{A^-_{\ell(M)}\} \subset \{A^-\}$.

(iii) $G = A^-_{\ell(M)}$ if and only if $\begin{cases} AGA = A \\ (AG)'M = MAG \end{cases}$

(k) Def. 2.2.3 A matrix denoted by A^+_{MN} $(=A^+)$ is said to be a minimum N-norm M-least squares inverse of A if

$x = A^+q$ is an M-least square solution of $Ax=y$ with a minimum N-norm, where M and N are p.d. matrices.

Remark 2.2.3

(i) A^+ is unique

(ii) if $G = A^+$ then following holds (and conversely)

$$AGA = A$$

$$GAG = G$$

$$(GA)'N = NGA \quad (N \text{ p.d.})$$

$$(AG)'M = MAG \quad (M \text{ p.d.})$$

2.3 Duality Theorem

Theorem 2.3.1

$$(X')^-_{M(V)} = [X^-_{\ell(V^{-1})}]'$$

Proof: Let $G = X^-_{\ell(V^{-1})}$.

Then

$$(2.3.1) \qquad G' = [X^-_{\ell(V^{-1})}]'$$

From the definition of G (Remarks 2.2.1 (iii)), we have

$$(XG)'V^{-1} = V^{-1}XG.$$

Therefore

$$XGV = V(XG)',$$

and

$$(G'X')'V = V(G'X').$$

Again by Remarks 2.2.1 (iii) we have

$$(2.3.2) \qquad G' = (X')^-_{M(V)}.$$

Combining (2.3.1) and (2.3.2) gives the result.

2.3.1 Application of Duality Theorem

(Another proof of the Gauss-Markoff Theorem).

Consider the following minimization problem. Minimize

$$(Y-X\beta)'\ V^{-1}\ (Y-X\beta).$$

A solution to the above problem is:

$$\hat{\beta} = X^{-}_{\ell(V^{-1})} Y \qquad (2.3.3)$$

Consequently an estimate of $p'\beta$ is $p'\hat{\beta} = p'X^{-}_{\ell(V^{-1})} Y$.

Next, suppose we want to find an estimate of $p'\beta$ by $L'Y$ such that

(a) $X'L = p$ (unbiasedness)

(b) $L'VL$ = minimum.

A solution to above problem is given by

$$L = (X')^{-}_{M(V)}\ p\ .$$

Thus an estimate of $p'\beta$ is

$$L'Y = p'[(X')^{-}_{M(V)}]'Y.$$

By the Duality Theorem, this solution can be written as

$$L'Y = p'[X^{-}_{\ell(V^{-1})}]Y.$$

From (2.3.3), the right hand side can be written as $p'\hat{\beta}$ which is a least squares solution.

2.4 Computation of A^- (Shows existence).

Let A be an nxn matrix of rank $r < n$.

(a) If A is symmetric then it has a spectral decomposition

$$A = \lambda_1 P_1 P_1' + \lambda_2 P_2 P_2' + \ldots + \lambda_r P_r P_r'$$

where $\lambda_1, \lambda_2 \ldots \lambda_r$ are non-zero eigenvalues of A with corresponding eigen-vectors $P_1, P_2, \ldots, P_r$. In such a case

$$A^- = \frac{1}{\lambda_1} P_1 P_1' + \frac{1}{\lambda_2} P_2 P_2' + \ldots + \frac{1}{\lambda_r} P_r P_r' \text{ with } P_i' P_j = 0 \;\; i \neq j.$$

(b) If A is not symmetric then it has a singular value decomposition (see [4] p. 38, [3] p. 42)

$$A = \lambda_1 P_1 Q_1' + \lambda_2 P_2 Q_2' + \ldots + \lambda_r P_r Q_r'$$

where $P_1, P_2, \ldots, P_r$ are the eigen vectors of AA' and
$Q_1, Q_2, \ldots, Q_r$ are the eigen vectors of A'A
λ_i are the positive square roots of the eigenvalues of A'A.

In this case

$$A^- = \frac{1}{\lambda_1} Q_1 P_1' + \ldots + \frac{1}{\lambda_r} Q_r P_r'$$

Remark 2.4.1 $P_1, P_2, \ldots, P_r$ are orthogonal to each other and $Q_1, Q_2, \ldots, Q_r$ are orthogonal to each other.

3.1 Condition of Consistency

Consider the GLM model

(3.1.1) $(Y, X\beta, \sigma^2 V)$.

It may be noted that the Gauss-Markoff model with restrictions on the parameter β

(3.1 $(Y, X\beta, \sigma^2 V); C = R\beta$

can be written as the GGM model

(3.1.3) $\quad (Y_e, X_e\beta, \sigma^2 V_e)$,

where

(3.1.4) $\quad Y_e = \binom{Y}{C}; \; X_e = \binom{X}{R}, \; V_e = \begin{pmatrix} V & 0 \\ 0 & 0 \end{pmatrix}$.

When V is singular in (3.1.1) there are some natural restrictions on the random vector Y and possibly on the parameter vector β.
One such restriction on Y is given by the following:

Lemma 3.1.1

$L'X = 0$, $L'V = 0$ implies that $L'Y = 0$ with probability 1.

Proof: The conditions

$E(L'Y) = L'X\beta = 0$

$Var(L'Y) = L'VL = 0$

imply that $L'Y = 0$ with probability 1. As a consequence of the above lemma, we have:

Theorem 3.1.1

(a) $Y \in \mathcal{M}(V:X)$ with probability 1.

This is called the consistency of the model.

(b) $R(V) = t < n$, implies the existence of an $(n \times s)$ matrix K such that $K'V = 0$. Here $s = (n-t)$ and the choice of $K = V^{\perp}$ works.

(c) $cov(K'Y) = \sigma^2 K'VK = 0$ implies that $K'Y = C$ (constant vector) with prob. 1.

(c) says there exist s independent linear functions of Y which are constants with probability 1.

Remark 3.1.1

Another way to state the above result is:

$Y - Y_0 \in \mathcal{M}(V)$ where Y_0 is an observed value of Y or $Y = Y_0 + VZ$ where Z is an arbitrary vector.

(a) Restrictions on the random variable Y

$K'Y = C.$

Therefore Y lies on the hyperplane $K'Y = C$.

We show that Y lies on a hyperplane through the origin.

Let $D = C^{\perp}$. Then

$D'K'Y = D'C = 0;$

i.e., $N'Y = 0$, where $N' = D'K'$. This implies that

$$Y \in \mathcal{M}(N^{\perp}).$$

(b) Restrictions on the parameter β

$E(K'Y) = K'X\beta = C.$

Therefore,

$$D'K'X\beta = 0 (D = C^{\perp}) \Rightarrow N'X\beta = 0$$

where $N' = D'K'$.

3.2 Unbiasedness of a Linear Estimator

Let us consider the model (1) and find the condition for a linear function $L'Y$ to be unbiased for $p'\beta$.

(3.2.1) $$E(L'Y) = L'X\beta = p'(\beta)$$

which must hold for all β such that

$$N'X\beta = 0$$

Then there exists a vector λ such that

$$L'X - p' = \lambda'N'X$$

or $$p = X'(L - N\lambda)$$

Thus we have the following lemmas.

Lemma 3.2.1

A necessary and sufficient condition that $p'\beta$ admits of a linear unbiased estimator is that $p \in \mathcal{M}(X')$.

Lemma 3.2.2

If $L'Y$ is unbiased for $p'\beta$ then it is necessary and sufficient that there exists a vector λ such that

(3.2.2) $$X'(L-N\lambda) = p$$

Lemma 3.2.3

If $L'Y$ is an unbiased estimator of $p'\beta$ then there exists a vector M such that

$$X'M = p$$

and $$L'Y = M'Y \text{ with prob. } 1.$$

Proof:

Take $M = L-N\lambda$ (λ as defined by (3.2.2)).

Then $M'Y = L'Y - \lambda'N'Y = L'Y$ with prob. 1. Also note that

$X'M = X'L - X'N\lambda = p$.

q.e.d.

Remark 3.2.1

(i) Note that when V is of full rank or when the observation Y is unknown, the condition for unbiasedness of $L'Y$ for $p'\beta$ is

$$X'L = p$$

which is usually given in textbooks.

This is not true in general as (3.2.2) shows.

(ii) Lemma (3.2.3) shows that the entire class of unbiased estimators of an estimable function $p'\beta$ can be generated by $M'Y$ where M satisfies the condition $X'M = p$.

Thus to find the minimum variance unbiased estimator of $p'\beta$ we need to determine M such that

$$M'VM \text{ is minimum}$$

subject to the condition $X'M = p$.

(iii) The result of Lemma 3.2.2 is based on the knowledge of the matrix N, which can be computed if V and a sample observation on the r.v. Y are known.

However if we want $L'Y$ to be unbiased for $p'\beta$ irrespective of the subspace to which Y may belong then the condition is

$$X'L = p.$$

Fortunately, in view of Lemma 3.2.3 the formulae we develop for the BLUE of $p'\beta$ and for the estimation of σ^2 are valid no matter which particular subspace Y may belong to.

4. THE IPM METHOD

4.1 Preliminaries

The Inverse Partition Matrix (IPM) Method requires the computation of a g-inverse of the partitioned matrix

$$(4.1.1) \qquad \begin{pmatrix} V & X \\ X' & 0 \end{pmatrix}^{-} = \begin{pmatrix} C_1 & C_2 \\ C_3 & -C_4 \end{pmatrix}$$

where V and X are defined as in the model (3.1.1). Once a g-inverse is computed by a suitable procedure we seem to have a Pandora's box supplying all the ingredients needed for obtaining the BLUE's, their variances and covariances, an unbiased estimate of σ^2, and test criteria without any further computations except for a few matrix multiplications. Thus the problem of inference from a linear model is reduced to the numerical problem of finding an inverse (a g-inverse) of the symmetric matrix given in (4.1.1).

We summarize some results about g-inverse of a partitioned matrix in the following.

Theorem 4.1.1

Let V, X, C_1, C_2, C_3, C_4, be as defined in (4.1.1). The following hold:

(i) $\begin{pmatrix} V & X \\ X' & 0 \end{pmatrix}^- = \begin{pmatrix} C_1' & C_3' \\ C_2' & -C_4' \end{pmatrix}$ is another choice of g-inverse.

(ii) $X\ C_3\ X = X = X\ C_2'\ X$

(iii) $V\ C_2\ X' = X\ C_2'\ V = X\ C_4\ X' = X\ C_4'\ X' = V\ C_3'\ X' = X\ C_3\ V$

(iv) $X'\ C_1\ X = 0,\ V\ C_1\ X = 0,\ X'\ C_1\ V = 0,\ V\ C_1\ V + X\ C_3\ V = V$

(v) $V\ C_1\ V\ C_1\ V = V\ C_1\ V = V\ C_1'\ V\ C_1\ V = V\ C_1'\ V$

(vi) Trace $V\ C_1 = R(V{:}X) - R(X)$

(vii) $\begin{pmatrix} C_1 \\ C_3 \end{pmatrix}$ is a g-inverse of $(V{:}X)$

(viii) $\begin{pmatrix} V & X \\ X' & 0 \end{pmatrix}^- = \begin{pmatrix} C_1 & C_3' \\ C_2' & -C_4 \end{pmatrix}$ is another choice of g-inverse.

Proof:

The result (i) is proved by taking transposes of either side of (4.1.1).

(ii) and (iii). We observe that the equations

(4.1.2) $\begin{cases} Va + Xb = 0 \\ X'a = X'd \end{cases}$

are solvable for any d, in which case

(4.1.3) $\begin{cases} a = C_2\ X'\ d \\ b = -C_4\ X'\ d \end{cases}$

is a solution.

Substituting (4.1.3) in (4.1.2) and omitting d, we have

(4.1.4) $$\begin{cases} V\, C_2\, X' = X\, C_4\, X' \\ X'\, C_2\, X' = X', \end{cases}$$

In view of (i) we can replace C_4, C_2 by C_4', C_3' in (4.1.4) to obtain

(4.1.5) $$\begin{cases} V\, C_3'\, X' = X\, C_4'\, X' \\ X'\, C_3'\, X' = X' . \end{cases}$$

Multiplying both sides of the first equation in (4.1.5) by $X\, C_3$, we obtain

(4.1.6) $$X\, C_3\, V\, C_3'\, X' = X\, C_3\, X\, C_4'\, X' = X\, C_4'\, X'.$$

So that $X\, C_4'\, X'$ is symmetric.

Then (4.1.4-6) prove the results (ii) and (i i).

(iv) We observe that the equations

(4.1.7) $$\begin{cases} Va + Xb = Xd, \\ X'a \qquad = 0, \end{cases}$$

are solvable for any d. Then,

(4.1.8) $$\begin{cases} a = C_1\, Xd, \\ b = C_3\, Xd, \end{cases}$$

is a solution. Substituting (4.1.8) in (4.1.7) and omitting d, we have

(4.1.9) $$\begin{aligned} V\, C_1\, X + X\, C_3\, X &= X, \\ X'\, C_1\, X \qquad\quad &= 0. \end{aligned}$$

But $X\, C_3\, X = X \Rightarrow V\, C_1\, X = 0$.

Also, $V\, C_1'\, X = 0$ in view of (i).

The result $V\, C_1\, V + X\, C_3\, V = V$ easily follows.

(V) We observe that the equations

(4.1.10) $$\begin{aligned} Va + Xb &= Vd, \\ X'a \qquad &= 0, \end{aligned}$$

are solvable for any d. One solution is

$$(4.1.11)\qquad \begin{aligned} a &= C_1\, Vd, \\ b &= C_3\, Vd. \end{aligned}$$

Substituting in (4.1.10) and omitting d

$$(4.1.12)\qquad \begin{aligned} V\, C_1\, V + X\, C_3\, V &= V, \\ X'\, C_1\, V \qquad\quad &= 0. \end{aligned}$$

This implies that

$$(4.1.13)\quad V\, C_1\, V\, C_1\, V + V\, C_1\, X\, C_3\, V = V\, C_1\, V = V\, C_1\, V\, C_1\, V,$$

since $V\, C_1\, X = 0$.

Also, since $V\, C_1'\, X = 0$,

$$(4.1.14)\qquad V\, C_1'\; V\, C_1\, V = V\, C_1'\, V,$$

and $V\, C_1\, V$ is symmetric.

(vi) To prove (vi) we use the result

$R(AA^-) = R(A) = Tr(AA^-)$ for any g-inverse A^- of A.

$$R\left[\begin{pmatrix} V & X \\ X' & 0 \end{pmatrix}\begin{pmatrix} C_1 & C_2 \\ C_3 & -C_4 \end{pmatrix}\right] = Tr\begin{pmatrix} V\, C_1 + X\, C_3 & V\, C_2 - X\, C_4 \\ X'\, C_1 & X'\, C_2 \end{pmatrix}$$

$$\begin{aligned} &= Tr(V\, C_1 + X\, C_3) + Tr\, X'\, C_2 \\ &= Tr(V\, C_1) + R(X\, C_3) + R(X'\, C_2) \end{aligned}$$

$$(4.1.15)\qquad = Tr(V\, C_1) + R(X) + R(X').$$

Moreover,

$$(4.1.16)\qquad R\begin{pmatrix} V & X \\ X' & 0 \end{pmatrix} = R(V{:}X) + R(X).$$

Equating (4.1.15) and (4.1.16) we have

$$Tr\, V\, C_1 = R(V{:}X) - R(X).$$

The results (vii) and (viii) are proved by direct verification.

Remark 4.1.1

(i) The results (ii), (iii) and (iv) of above theorem are necessary and sufficient for relation (4.1.1) to hold.

(ii) C_2 and C_3' are in fact minimum V-norm g-inverse of X'.

(iii) $(C_1':C_3')$ is a minimum V-norm g-inverse of $\binom{V}{X'}$.

As remarked earlier the inverse matrix (4.1.1) is like a Pandora's Box which gives all that is necessary for drawing inference on the β-parameters. We state the results in Theorem 4.2.1 which demonstrates the use of the sub-matrices in (4.1.1).

4.2 Main Results

Theorem 4.2.1

Let C_1, C_2, C_3, C_4 be as defined in (4.1.1). Then the following hold:

(i) [Use of C_2 or C_3]. The BLUE of an estimable parametric function $p'\beta$ is $p'\hat{\beta}$ where

(4.2.1) $$\hat{\beta} = C_2' Y \text{ or } \hat{\beta} = C_3 Y$$

(ii) [Use of C_4]. The dispersion matrix of $\hat{\beta}$ is $\sigma^2 C_4$ in the sense,

(4.2.2) $$\mathrm{Var}(p'\hat{\beta}) = \sigma^2 p' C_4 p,$$

(4.2.3) $$\mathrm{Cov}(p'\hat{\beta}, q'\hat{\beta}) = \sigma^2 p' C_4 q = \sigma^2 q' C_4 p,$$

where $p'\beta$ and $q'\beta$ are estimable.

(iii) [Use of C_1]. An unbiased estimator of σ^2 is

(4.2.4) $$\hat{\sigma}^2 = f^{-1} Y' C_1 Y$$

where

$$f = R(V:X) - R(X)$$

Proof:

(i) If $L'Y$ is an unbiased estimator of $p'\beta$ then

$$X'L = p.$$

Subject to this condition

$$V(L'Y) = \sigma^2 L'VL,$$

or $L'VL$ has to be minimized to obtain the BLUE of $p'\beta$.

Let L_* be an optimum choice and L be any other vector such that

$$X'L = X'L_*.$$

Then

$$L'VL = (L-L_* + L_*)' V(L-L_* + L_*)$$

$$= (L-L_*)' V(L-L_*) + L_*'VL_* + 2L_*'V(L-L_*) \geq L_*'VL_*,$$

iff $L_*'V(L-L_*) = 0$ whenever $X'(L-L_*) = 0$; i.e., $VL_* = -XK_*$ for a suitable K_*.

Then L_* and K_* satisfy the equations

$$(4.2.5) \qquad \begin{cases} VL_* + XK_* = 0 \\ X'L_* \qquad\quad = p \end{cases}$$

We observe that the equations (4.2.5) admit a solution and any two solutions L_{1*} and L_{2*} satisfy the condition

$$V(L_{1*} - L_{2*}) = 0.$$

Since (4.2.5) is consistent, a solution is given by

$$\begin{matrix} L_* = C_2p \\ K_* = -C_4p \end{matrix} \quad \text{or} \quad \begin{matrix} L_* = C_3'p \\ K_* = -C_4p \end{matrix}$$

Then the BLUE of $p'\hat{\beta}$ is

$$L_*'Y = p'C_2'Y = p'C_3Y.$$

(ii) We use the fact that $p = X'M$ for some M. Then

$$\begin{aligned}
\mathrm{Var}(p'C_2'Y) &= \sigma^2 M'(XC_2'V)C_2X'M \\
&= \sigma^2 M'XC_4(X'C_2X')M && \text{Using Theorem 4.1.1 (iii)} \\
&= \sigma^2 M'XC_4X'M && \text{Using Theorem 4.1.1 (ii)} \\
&= \sigma^2 p'C_4p.
\end{aligned}$$

Similarly,

$$\mathrm{Cov}(p'C_2'Yq'C_2'Y) = \sigma^2 p'C_4q = \sigma^2 q'C_4p.$$

(iii) Since $X'C_1V = 0$ and $X'C_1X = 0$, using Theorem 4.1.1 (iv),

$$Y'C_1Y = (Y - X\beta)'\ C_1(Y-X\beta).$$

We have

$$\begin{aligned}
E[(Y-X\beta)'C_1(Y-X\beta)] &= \sigma^2\ \mathrm{Tr}\ C_1[E\{(Y-X\beta)\ (Y-X\beta)'\}] \\
&= \sigma^2\ \mathrm{Tr}\ C_1V = \sigma^2[R(V{:}X) - R(X)],
\end{aligned}$$

where the last equality follows from (vi) of Theorem 4.1.1.

Theorem 4.2.2

Let $P'\hat{\beta}$ be the vector of BLUE's of a set of k estimable parametric functions $P'\beta$, $R_0^2 = Y'C_1Y$ and f be as defined in Theorem 4.2.1. If $Y \sim N_n(X\beta,\sigma^2 V)$, then:

(i) $P'\hat{\beta}$ and $Y'C_1Y$ are independently distributed with

(4.2.6) $$P'\hat{\beta} \sim N_k(P'\beta,\ \sigma^2 D)$$

and

(4.2.7) $$Y'C_1Y \sim \sigma^2 \chi_f^2$$

where $D = P'C_4P$.

(ii) Let $P'\beta = w$ be the null hypothesis. The null hypothesis is consistent iff

(4.2.8) $$DD^-u = u$$

where $u = P'\hat{\beta} - w$.

If the hypothesis is consistent, then

(4.2.9) $$F = \frac{u'D^{-}u}{h} \div \frac{R_0^2}{f}, \qquad h = R(D)$$

has a Central F distribution on h and f degrees of freedom when the hypothesis is true, and a non-central F distribution when the hypothesis is false.

Proof (i)

The result (4.2.6) is easy to establish.

(4.2.7) follows since

$$Y'\left(\frac{C_1 + C_1'}{2}\right)Y = (Y-X\beta)'\left(\frac{C_1 + C_1'}{2}\right)(Y-X\beta)$$

and by

$$VC_1VC_1V = VC_1V$$

and

$$VC_1'VC_1V = VC_1V$$

which is an NAS condition for a χ^2 - dist. (See Rao [3] p. 188 and also Rao and Mitra [4].)

The degrees of freedom of the χ^2 is

$$Tr\ VC_1 = R(V:X) - R(X) = f,$$ using Theorem 4.1.1, result (vi).

Since $P'\beta$ is estimable,

$$P' = QX \text{ for some } Q.$$

Then $P'\hat{\beta} = QXC_3Y$.

The condition for independence of $Y'\left(\frac{C_1 + C_1'}{2}\right)Y$ and QXC_3Y is

$$V\left(\frac{C_1 + C_1'}{2}\right)VC_3'X'Q'QXC_3V = 0$$

which is true since

$$V\left(\frac{C_1 + C_1'}{2}\right)VC_3'X' = V\left(\frac{C_1 + C_1'}{2}\right)XC_3V = 0$$

Using Theorem 4.1.1 (ii) and (iii).

(ii) The hypothesis $P'\beta = w$ is consistent for any vector M.

$\mathrm{Var}[M'(p'\hat{\beta}-w)] = 0 \Rightarrow M'(P'\hat{\beta}-w) = 0$,

i.e., $M'DM = 0 \Rightarrow M'u = 0$ or $u \in \mathcal{M}(D)$, for which a NAS condition

is $DD^-u = u$, for any g-inverse D^- of D.

Since dispersion matrix of $u = \sigma^2 D$ and

$$DD^-D = D,$$

$$\frac{u'D^-u}{\sigma^2} \sim \chi^2_h \qquad h = R(D),$$

Using the result proved in (i), R_0^2 is distributed as χ^2_f

independently of u. Hence the result (4.2.9) follows.

Q.E.D.

In Theorem 4.2.2 the numerator of the F statistic for testing the linear hypothesis $p'\beta = w$ was obtained in the form $U'D^-U$ which involved the estimation of deviations in individual hypotheses, computations of their dispersion matrix and its inverse.

Theorem 4.2.3 provides an alternative method of computing the numerator as in the theory of least squares.

Theorem 4.2.3

Let C_1 be as defined in (4.1.1) and

$$\begin{pmatrix} V & 0 & X \\ 0 & 0 & P' \\ X' & P & 0 \end{pmatrix}^- = \begin{pmatrix} E_1 & E_2 \\ E_3 & E_4 \end{pmatrix}$$

for any choice of the g-inverse. Further let Y have a MVN distribution. Then the hypothesis $P'\beta = w$ is consistent iff

$$\begin{pmatrix} V & 0 & X \\ 0 & 0 & P' \end{pmatrix} \begin{pmatrix} E_1 \\ E_3 \end{pmatrix} \begin{pmatrix} Y \\ W \end{pmatrix} = \begin{pmatrix} Y \\ W \end{pmatrix}$$

in which case

$$T_1 = \begin{pmatrix} Y \\ W \end{pmatrix}' E_1 \begin{pmatrix} Y \\ W \end{pmatrix} - Y'C_1Y \quad \text{and}$$

$$T_2 = Y'C_1Y$$

are independently distributed as $\sigma^2\chi^2$ on

$$d = R\begin{pmatrix} V & X \\ 0 & P' \end{pmatrix} - R(V:X) \quad \text{and}$$

$$f = R(V:X) - R(X)$$

degrees of freedom respectively.

Hence,

$$F = \frac{T_1}{d} \div \frac{T_2}{f}$$

has the F distribution on d and f degrees of freedom.

5. UNIFIED LEAST SQUARES METHOD (ULS)

5.1 Statement of the Problem

Suppose we have a GGM model

(5.1.1) $$(Y, X\beta, \sigma^2 V).$$

(a) When $V = I$ and X is of full rank in (5.1.1) Gauss [2] propounded the famous theory of least squares which postulates that the best estimate $\hat{\beta}$ of β is obtained by minimizing the sum of squares

(5.1.2) $$(Y-X\beta)' (Y-X\beta)$$

Gauss showed in fact that $\hat{\beta}$ is the BLUE of β and that an unbiased estimator of ^{2} is

(5.1.3) $$\hat{\sigma}^2 = (Y-X\hat{\beta})' (Y-X\hat{\beta})/n-r \quad \text{with}$$

$$r = R(X).$$

(b) If $V \neq I$ but non-singular we make the transformations

$$Y_t = V^{-1/2} Y$$
(5.1.4)
$$X_t = V^{-1/2} X$$

and reduce the problem to the original Gaussian model

(5.1.5) $$(Y_t, X_t\beta\ \sigma^2 I)$$

to which the theory of least squares is applicable.

Thus we are led to minimize

(5.1.6) $$(Y_t - X_t\beta)'\ (Y_t - X_t\beta) = (Y - X\beta)'\ V^{-1}\ (Y - X\beta)$$

which is the procedure proposed by Aitken [1].

(c) If V is singular, Aitken's procedure fails as V^{-1} does not exist. (e.g. if V is symmetric and n.n.d. with $R(V) = r < n$).

In such a case V has the following spectral decomposition

$$V = \lambda_1 P_1 P_1' + \ldots + \lambda_r P_r P_r' .$$

Let $P_{r+1},\ldots,P_n$ denote eigenvectors corresponding to 0 eigenvectors.

Suppose $F_i = \lambda_i^{-1/2} P_i$ $i = 1,2,\ldots,r$.

Then $\mathrm{Var}(F_i'Y) = \sigma^2 F_i'VF_i = \sigma^2\lambda_i^{-1}P_i'VP_i = \sigma^2$.

$$\mathrm{Cov}(F_i'Y, F_j'Y) = \sigma^2 P_i'VP_j = 0.$$

Also let $B_j = P_{r+j}$ $j = 1,2,\ldots,n-r$, then

$$\mathrm{Var}(B_j'Y) = B_j'VB_j = 0.$$

Letting

$$F = (F_1, F_2,\ldots,F_r),$$

$$B = (B_1, B_2,\ldots,B_{n-r}),$$

the given model $(Y, X\beta, \sigma^2 V)$ is reduced to $(F'Y, F'X\beta, \sigma^2 I_r)$ with constraints $B'X\beta = B'Y = C$.

This is a simple Gauss model with restrictions on the parameter. Hence the original problem is reduced to

$$\begin{cases} \text{minimize:} & (F'Y - F'X\beta)'\,(F'Y - F'X\beta), \\ \text{such that} & B'X\beta = C\,; \end{cases}$$

or

$$(5.1.7)\quad \begin{cases} \text{minimize} & (Y-X\beta)'\,FF'\,(Y-X\beta), \\ \text{such that} & B'X\beta = C. \end{cases}$$

Since $F = (\lambda_1^{-1/2}\, P_1, \ldots, \lambda_r^{-1/2}\, P_r)$,

$$FF' = \lambda_1^{-1}\, P_1 P_1' + \ldots + \lambda_r^{-1}\, P_r\, P_r' = V^- \,.$$

Therefore FF' can be identified with V^- and (5.1.7) can be reformulated as

$$(5.1.8)\quad \begin{cases} \text{minimize:} & (Y-X\beta)'\,V^-\,(Y-X\beta), \\ \text{such that} & B'X\beta = C. \end{cases}$$

<u>Remark 5.1.1</u>

Here V^- can be taken as <u>any</u> g-inverse of V. The solution to (5.1.8) is obtained by solving the equation

$$(5.1.9)\quad \begin{cases} X'V^-\,X\beta + X'\,B\lambda = X'V^-\,Y \\ B'X\beta \qquad\qquad\qquad = C. \end{cases}$$

Let

$$\begin{pmatrix} X'V^-\,X & X'B \\ B'X & 0 \end{pmatrix}^- = \begin{pmatrix} H_1 & H_2 \\ H_3 & H_4 \end{pmatrix}.$$

Then the solution of (5.1.9) is

$$\hat{\beta} = H_1 X'V^-\,Y + H_2 C$$

and

$$\mathrm{Var}(p'\hat{\beta}) = \sigma^2 p' H_1 X'V^-\,VV^-\,XH_1' p.$$

Also, as before we can calculate

(1) $R_0^2 = \min_{B'X\beta = C} (Y-X\beta)' V^-(Y-X\beta)$,

(2) $R_1^2 = \min_{\substack{B'X\beta = C \\ P'\beta = W}} (Y-X\beta)' V^-(Y-X\beta)$.

(3) The F statistic (4.2.9) for testing

H_0: $P'\beta = W$ is

$$\frac{R_1^2-R_0^2}{h} \div \frac{R_0^2}{f} .$$

Motivated by the knowledge that Aitken's procedure fails when V is singular, we raise the following question. Does there exist a matrix M regardless of whether V is singular or not such that the following conditions hold?

(a) The BLUE of any estimable ($p \in \mathcal{M}(X')$) parametric function $p'\beta$ is $p'\hat{\beta}$ where $\hat{\beta}$ is a stationery point of the function

(5.1.10) $$(Y-X\beta)' M(Y-X\beta)$$

i.e. where the derivative of (5.1.10) with respect to β vanishes is zero.

(b) An unbiased estimator of σ^2 is obtained as

(5.1.11) $$\hat{\sigma}^2 = (Y-X\hat{\beta})' M(Y-X\hat{\beta}) \div f$$

where

$$f = R(V:X) - R(X).$$

(c) R_1^2 = stationary value of

(5.1.12) $$(Y-X\beta)' M(Y-X\beta)$$

under the restriction $P'\beta = W$,

R_0^2 = stationary value of

(5.1.13) $$(Y-X\beta)' M(Y-X\beta)$$

and

$$F = \frac{R_1^2 - R_0^2}{h} \div \frac{R_0^2}{f} .$$

Remark 5.1.2

It can be assumed, M may be chosen to be symmetric. Theorem 5.2.1 and [illegible] 2 provide complete answers to the questions (a) and (b). We show that [illegible] singular or not, the choice should be

$$(5.1.14) \qquad M = (V + XUX')^{-}$$

for any symmetric g-inverse where U is any symmetric matrix such that

$$\mathcal{M}(V:X) = \mathcal{M}(V + XUX').$$

In particular we can always choose

$$(5.1.15) \qquad M = (V + k^2XX')^{-}$$

for any choice of g-inverse where k is an arbitrary non-zero constant. (c) cannot hold for any choice of M for all testable hypotheses.

5.2 Some Preliminary Lemmas

Lemma 5.2.1 Let T be a matrix such that $R(X'TX) = R(X)$. Then

$$(5.2.1) \qquad X(X'TX)^{-}(X'TX) = X$$

Proof:

If $R(X'TX) = R(X)$, then for any vector λ,

$$X'TX\lambda = 0 \text{ if and only if } X\lambda = 0.$$

This result together with the identity

$$(5.2.2) \qquad 0 = X'TX(X'TX)^{-}X'TX - X'TX = X'TX[(X'TX)^{-}X'TX - I],$$

yields (5.2.1).

Lemma 5.2.2 Let U be symmetric and V be n.n.d. matrices such that

$$(5.2.3) \qquad \mathcal{M}(V:X) = \mathcal{M}(V + XUX')$$

Then

$$R[X'(V + XUX')^{-}X] = R(X')$$

Proof:

The result is easy to establish using Lemma 5.2.1.

<u>Lemma 5 .3</u> Let $X\hat{\beta}$ be the BLUE of $X\beta$. Then the unbiased estimat... of ... is

(5.2.4) $\quad f^{-1}(Y-X\hat{\beta})'V^{-}(Y-X\hat{\beta}) = f^{-1}(Y-X\hat{\beta})'(V + XUX')^{-}(Y-X\beta)$

where

$$f = R(V:X) - R(X) \text{ and } U \text{ is defined in Lemma 5.2.2.}$$

<u>Proof:</u>

The left-hand side expression in (5.2.4) is well-known.and the equivalenc... with the right-hand side follows easily observing that $Y-X\hat{\beta} \in \mathcal{M}(V)$.

<u>Theorem 5.2.1</u>

Let $(Y,X\beta,\sigma^2V)$ be a GGM model and M be a symmetric matrix such that

$$\mathcal{M}(X'MV) \subset \mathcal{M}(X'MX)$$

in which case

$$(Y-X\beta)'M(Y-X\beta)$$

as a function of β has stationary values. Further let $\hat{\beta}$ be a stat... ...nt. If $p'\hat{\beta}$ is the BLUE of $p'\beta$ for every $p\in\mathcal{M}(X')$. Then it is necessary ...

(5.2.5) $$R(X'MX) = R(X)$$

and M is of the form

(5.2.6) $$(G + XUX')^{-} + K$$

for any symmetric choice of g-inverse where U and K are any symmetric ma...ces such that

(5.2.7) $$\mathcal{M}(V:X) = \mathcal{M}(V + XUX')$$

(5.2.8) $$VKX = 0 \quad X'KX = 0$$

<u>Conversely:</u> If M is of the form (5.2.6) with (5.2.7) and (5.2.8) true, ...en

$$R(X'MX) = R(X) \text{ and } p'\hat{\beta} \text{ is the BLUE of } p'\beta$$

for every $p\in\mathcal{M}(X')$.

<u>Proof:</u>

Equating the derivative of $(Y-X\beta)'M(Y-X\beta)$ to zero, we obtain

(5.2.9) $$X'MX\beta = X'MY$$

which is consistent since $\mathcal{M}(X'MX) \supset \mathcal{M}(X'MV)$ and $Y \in \mathcal{M}(V:X)$ with probabili., 1.

In this case

(5.2.10) $$\hat{\beta} = (X'MX)^{-} X'MY$$

is a stationary point.

Let $p = X'L$. Then $p'\hat{\beta}$ is the BLUE of $p'\beta$ and it follows by definition ...

(5.2.11) $$L'X(X'MX)^{-} X'MX = L'X.$$

Since L is arbitrary in (5.2.11), we have

$$X(X'MX)^{-} (X'MX) = X \Rightarrow R(X'MX) = R(X),$$

which proves (5.2.5) .

If $p'\hat{\beta}$ is the BLUE of $p'\beta$ for every $p \in \mathcal{M}(X')$ then applying the lemma on p. 17 of Rao [3], we have

(5.2.12) $$L'X(X'MX)^{-} X'MVZ = 0 \text{ for any } L,$$

where Z is a matrix of maximum rank such that

$$X'Z = 0.$$

Then (5.2.12) $\Rightarrow X(X'MX)^{-} X'MVZ = 0 \Leftrightarrow X'MVZ = 0$

(5.2.13) $$\Rightarrow VMX = XQ$$

for some Q.

Now there exists a symmetric matrix U such that

(5.2.14) $$X'M(V + XUX')MX = X'MX.$$

Let $W = X'MX$. Then it can be verified easily that one choice of U is

$$W^{-} (-X'MVMX + W)W^{-},$$

where W^{-} is a symmetric g-inverse of W.

Multiplying both sides of (5.2.14) by $X(X'MX)^{-}$ and using (5.2.13) and Lemma 5.2.1, we obtain

(5.2.15) $$(V + XUX')MX = X.$$

If $\alpha'(V + XUX') = 0$ then from (5.2.15) $\alpha'X = 0$ and hence $\alpha'V = 0$ and vice-versa, which proves (5.2.7).Choosing a symmetric g-inverse and a symmetric K, let

(5.2.16) $$M = (V - XUX')^{-} + K.$$

Substituting (5.2.16) in (5.2.15), we obtain (5.2.8).

The converse is easy to prove using Lemma 5.2.2.

Theorem 5.2.2

Let $\hat{\beta}$ be a stationary point of $(Y-X\beta)'M(Y-X\beta)$ where M is a symmetric matrix such that

$$\mathcal{M}(X'MX) \supset \mathcal{M}(X'MV).$$

If $p'\hat{\beta}$ is the BLUE of $p'\beta$ for every $p \in \mathcal{M}(X')$ and for all $Y \in \mathcal{M}(V:X)$,

(5.2.17) $$\hat{\sigma}^2 = f^{-1}(Y-X\hat{\beta})'M(Y-X\hat{\beta})$$

is an unbiased estimator for σ^2, then it is necessary and sufficient that M is a symmetric g-inverse of $V + XUX'$ where U is any symmetric matrix such that $\mathcal{M}(V:X) = \mathcal{M}(V + XUX')$.

Proof:

We have already seen that M is of the form (5.2.6) and K satisfies (5.2.8).
If (5.2.17) is the same as (5.2.4), then

$$(Y-X\hat{\beta})'K(Y-X\hat{\beta}) = 0 = Y'KY.$$

ing (5.2.8) for all $Y \in \mathcal{M}(V + XUX')$, which implies that $VKV = 0$ in addition to (5.2.8). Then,

$$(V + XUX')[(V + XUX')^{-} + K](V + XUX') = (V + XUX')$$

which shows that M is a g-inverse of $(V + XUX')$.

Remark 5.2.1

(i) In Theorem 5.2.1 we showed that M is a symmetric g-inverse of $(V + XUX')$. It may be seen that the expression

$$(Y-X\beta)'(V + XUX')^{-} (Y-X\beta)$$

is independent of the choice of a g-inverse and in practice one can use any g-inverse.

(ii) If $V + XUX'$ is a n.n.d. matrix then

$$(Y-X\beta)'(V + XUX')^{-} (Y-X\beta)$$

is independent of the choice of the g-inverse, is non-negative and obtains a minimum at $\hat{\beta}$ where the derivative vanishes.

We can always choose U in such a way that $V + XUX'$ is n.n.d. and satisfies (5.2.7). For example U can be any p.d. matrix.

(iii) It may be seen that $(V + XUX')^{-}$ need not be a g-inverse of V.

If there exist a matrix U such that $V + XUX'$ satisfies (5.2.7) and

$$\{(V + XUX')^{-}\} \subset \{V^{-}\}$$

then it can be shown that a NAS condition is:

$$\mathcal{M}(V) \cap \mathcal{M}(XUX') = \{0\}.$$

Such a choice of U can be made if necessary.

Theorem 5.2.3

Let M be chosen as in Theorem 5.2.2 and $P'\beta$ be a set of k estimable functions i.e

$$\mathcal{M}(P) \subset \mathcal{M}(X').$$

Then $P'\hat{\beta}$ are the BLUE's of $P'\beta$ and the dispersion matrix of $P'\hat{\beta}$ is

(5.2.18) $$D(P'\hat{\beta}) = \sigma^2 P'[(X'(V + XUX')^{-} X)^{-} - U]P$$

Proof:

Let $W = (V + XUX')$.

Then

$$P'\hat{\beta} = P'(X'W^{-} X)^{-} X'W^{-}Y$$

and

(5.2.19) $$D(P'\hat{\beta}) = \sigma^2 P'(X'W^{-} X)^{-} X'W^{-} V[P'(X'W^{-} X)^{-} X'W^{-}]'$$

Write $V = V + XUX' - XUX'$

$= W - XUX'$ in (5.2.19).

Then, repeatedly using the relation (a) of Section 2.2 we get (5.2.18). Finally, to test the hypothesis

$$P'\beta = W.$$

We proceed as follows:

Let $u = P'\hat{\beta} - W$ and

$$D(u) = \sigma^2 D, \ R(D) = h.$$

The hypothesis is consistent if

(5.2.20) $$DD^- u = u.$$

If (5.2.20) holds, then the null distribution of the statistic

(5.2.21) $$F = \frac{u'D^- u}{h} \div \hat{\sigma}^2$$

is the F distribution on h and f degrees of freedom when $Y \sim MVN(X\beta, \sigma^2 V)$. The results (5.2.20) and (5.2.21) are proved in section 4.

In Theorems 5.2.1 and 5.2.2, it is shown that there exists a matrix M, whether V is nonsingular or not, such that a stationary value $\hat{\beta}$ of

$$(Y-X\beta)'M(Y-X\beta)$$

provides the BLUE of an estimable function $p'\beta$ as $p'\hat{\beta}$, and an unbiased estimator of σ^2 is

$$\hat{\sigma}^2 = f^{-1}(Y-X\hat{\beta})'M(Y-X\hat{\beta}).$$

So far we have an analogue of the least squares theory in the general case.

The first departure from the least squares results is the expression (5.2.18) for the dispersion matrix of $P'\hat{\beta}$, which contains the extra term $\sigma^2 P'UP$. It can be shown that there exists no choice of M unless $\mathcal{M}(X) \subset \mathcal{M}(V)$ such that

(5.2.22) $$D(P'\hat{\beta}) = \sigma^2 P'(X'MX)^- P$$

for all P such that $P'\beta$ is estimable.

Since (5.2.22) does not hold, there exists no choice of M, unless $\mathcal{M}(x) \subset \mathcal{M}(V)$, which enables the computation of the numerator of the F-statistic, (5.2.21), in the form

$$u'D^{-}u = \min_{P'\beta=w} (Y-X\beta)'M(Y-X\beta) - \min(Y-X\beta)'M(Y-X\beta)$$

for all testable hypotheses of the form $P'\beta = w$.

However Rao [6] and Mitra [10] have shown that a suitable choice of M can be made provided the null hypothesis is written in a modified but an equivalent form. The computation of such an M is somewhat complicated and it is much simpler to compute the F-statistic as in Theorem 5.2.3, using the simple choice of M as in Theorems 5.2.1 and 5.2.2 for estimating $P'\beta-w$ and σ^2. Note that M can always be chosen as $(V+XX')^{-}$, which satisfies the conditions of the Theorem 5.2.2 (see Rao and Mitra [12]).

6. BLUE'S AS PROJECTIONS

6.1 Projection Operators

It is well known that when V is nonsingular the BLUE of $X\beta$ is obtained by the orthogonal projection of Y on $\mathcal{M}(x)$, using the norm $||x|| = (x'V^{-1}x)^{1/2}$, which is the same as the projection of Y on $\mathcal{M}(x)$ along $\mathcal{M}(VZ)$, where $Z = X^{\perp}$. [Note that $\mathcal{M}(x)$ and $\mathcal{M}(VZ)$ are disjoint subspaces whether V is nonsingular or not]. We prove the corresponding results when V is singular. Naturally, the results have to be stated in a slightly different manner since V^{-1} does not exist (hence the norm $||x||$ cannot be defined as in the nonsingular case), and $\mathcal{M}(x)$ and $\mathcal{M}(VZ)$, although disjoint, may not span the entire space E^n (hence the projection on $\mathcal{M}(x)$ along $\mathcal{M}(VZ)$ is not properly defined).

Definition 6.1.1. Let Σ be an n.n.d. (non-negative definite) matrix of order n and define Σ-norm as

(6.1.1) $$||x||_{\Sigma} = (x'\Sigma x)^{1/2}.$$

Further let A be an nxm matrix. We call $P_{A\Sigma}$ a projector into $\mathcal{M}(A)$ under the Σ-norm if

$$(6.1.2) \quad \begin{cases} \mathcal{M}(P_{A\Sigma}) \subset \mathcal{M}(A) \\ \\ \|y-P_{A\Sigma}\|_\Sigma \leq \|y - A\lambda\|_\Sigma \text{ for all } y\in E^n,\ \lambda\in E^m. \end{cases}$$

The following lemma is easily established (see Mitra and Rao, [11]).

Lemma 6.1.1. If $P_{A\Sigma}$ is as defined in (6.1.2), it is necessary and sufficient that

$$(6.1.3) \qquad \mathcal{M}(P_{A\Sigma}) \subset \mathcal{M}(A)$$

$$(6.1.4) \qquad (P_{A\Sigma})'\Sigma P_{A\Sigma} = \Sigma P_{A\Sigma} = (P_{A\Sigma})'\Sigma$$

$$(6.1.5) \qquad \Sigma P_{A\Sigma}A = \Sigma A$$

Definition 6.1.2. Let U and W be two matrices such that $\mathcal{M}(U)$ and $\mathcal{M}(W)$ are disjoint, which together may not span the entire space. Any vector $\alpha\in\mathcal{M}(U:W)$ has the unique decomposition

$$\alpha = \alpha_1 + \alpha_2;\ \alpha_1 \in \mathcal{M}(U),\ \alpha_2 \in \mathcal{M}(W).$$

Then $P_{U|W}$ is said to be a projector onto $\mathcal{M}(U)$ along $\mathcal{M}(W)$ iff

$$(6.1.6) \qquad P_{U|W}\ \alpha = \alpha_1 \quad \text{for all } \alpha \in \mathcal{M}(U:W).$$

The following lemma is easily established.

Lemma 6.1.2. If $P_{U|W}$ is a projector as given in Definition 2, then it is necessary and sufficient that

$$(6.1.7) \qquad P_{U|W}\ U = U,\ P_{U|W}\ W = 0$$

and one choice of $P_{U|W}$ is

$$(6.1.8) \qquad P_{U|W} = U(GU)^{-}G,$$

where $G' = W^{\perp}$ and $(GU)^{-}$ is any g-inverse of GU.

6.2 Applications

The following theorem provides expressions for the BLUE's in terms of projection operators as described in definitions 1 and 2.

Theorem 6.2.1. Consider the G.G.M. model $(Y, X\beta, \sigma^2 V)$. Then the following hold:

(i) Let $L'Y$ be an unbiased estimator of $p'\beta$ with the property $L'X = p'$, and define $L_* = (I-P_{ZV})L$, where $Z = X^\perp$. Then $L_*'Y$ is the BLUE of $p'\beta$.

(ii) Let $S = V+XX'$, S^- be any n.n.d. g-inverse of S, and $Z = X^\perp$. Then

(6.2.1) $$(P_{ZV}' + P_{XS^-})\, a = a \quad \text{for any } a \in \mathcal{M}(V:X)$$

i.e., the sum of the projection operators on the left hand side of (6.2.1) is an identity in the space $\mathcal{M}(V:X) = \mathcal{M}(VZ:X)$.

(iii) The BLUE of $X\beta$ is

(6.2.2) $$(I-P_{ZV}')Y = (P_{XS^-})Y = (P_{X|VZ})Y$$

where the projection operators are as described in definitions 1 and 2.

Proof of (i). Since $\mathcal{M}(P_{ZV}) \subset \mathcal{M}(Z)$, $P_{ZV}'X = 0$ and hence

$$E(L'P_{ZV}'Y) = L' \cdot P_{ZV}'X = 0,$$

giving

$$E(L_*'Y) = E(L'Y) - E(L'P_{ZV}'Y) = E(L'Y)$$

so that $L_*'Y$ is unbiased for $p'\beta$. Further

$$L_*'VZ = L'(I-P_{ZV})'VZ = 0$$

using the conditions (7.4) and (7.5), which shows that $L_*'Y$ has minimum variance.

Proof of (ii). Since $\mathcal{M}(V:X) = \mathcal{M}(VZ:X)$ we need only verify that

$$(P_{ZV}' + P_{XS^-})\,(VZ:X) = (VZ:X)$$

which follows from the definitions of the projection operators.

Proof of (iii). From (i) it follows that the BLUE of $X\beta$ is

$$(I-P'_{ZV})Y$$

and from (ii) we have

$$(I-P'_{ZV})Y = (P_{XS^-})Y$$

To prove the last part of the equality in (6.2.2), consider the unique decomposition

(6.2.3) $$Y = XY_1 + VZY_2$$

on the disjoint subspaces $\mathcal{M}(x)$ and $\mathcal{M}(VZ)$. Note that $XY_1 = (P_{X|VZ})Y$ where $P_{X|VZ}$ is the projector onto $\mathcal{M}(x)$ along $\mathcal{M}(VZ)$. Now

$$Xp = E(Y) = XE(Y_1) + VZE(Y_2),$$

$$\Rightarrow x[\beta-E(Y_1)] = VZE(Y_2) = 0$$

since $\mathcal{M}(x)$ and $\mathcal{M}(VZ)$ are disjoint. Then $E(XY_1) = E(Y)$, so that XY_1 is unbiased for $x\beta$.

Further from (6.2.3)

$$Cov(Y,Z'Y) = X\ Cov(Y_1,Z'Y) + VZ\ Cov(Y_2,Z'Y)$$

(6.2.4) $$\Rightarrow VZ = XD_1 + VZD_2 \quad \text{for some } D_1 \text{ and } D_2$$

$$\Rightarrow VZ(I-D_2) = XD_1 = 0 = Cov\ (XY_1,Z'Y)$$

which shows that XY_1 is the BLUE of $E(XY_1) = x\beta$.

Theorem 6.2.1 is thus completely proved.

Note that, following (6.1.8), we can represent

(6.2.5) $$P_{X|VZ} = X(GX)^-G$$

where $G' = (VZ)^{\perp}$. When V=I, we have $G = X'$ giving the BLUE of $x\beta$ as

(6.2.6) $$(P_{X|VZ})Y = X(X'X)^-X'Y.$$

When V is nonsingular, we have $G = X'V^{-1}$ giving the BLUE of $X\beta$ as

(6.2.7) $$P_{X|VZ}\, Y = X(X'V^{-1}X)^{-}X'V^{-1}Y.$$

Thus (6.2.5) provides the well known formulae (6.2.6) and (6.2.7) in the particular cases considered.

In these lectures we have considered the problem of estimating $p'\beta$ by $L'Y$ such that $L'VL$ is a minimum subject to $X'L = p$, which provides a complete solution to the BLUE. However, this approach does not provide all possible representations of the BLUE. For this, one has to minimize $L'VL$ subject to the condition $X'L-p \in \mathcal{M}(X'N)$ where N is as defined in Section 3.1. The latter problem called BLUE(W), BLUE in wider sense, which is of some theoretical interest is solved in Rao [9].

Note. The references given at the end of the notes constitute the material on which the lectures were based. For references to related work by other authors the reader is referred to bibliographies in Rao [5] and [6]. It may be noted that Goldman and Zelen [13] were the first to consider the case of nonsingular V in a systematic way using generalized inverses.

References

1. Aitken, A. C. (1934). On Least Squares and Linear Combinations of Observations. Proc. Roy. Soc. Edinb. A, 55, 42-47.

2. Gauss, C. F. (1855). Methode des Moindres Carres (Trans).

3. Rao, C. R. (1973). Linear Statistical Inference, Wiley, New York. (Second edition)

4. Rao, C. R. and Mitra, S. K. (1971). Generalized Inverse of a Matrix and its Applications, Wiley, New York.

5. Rao, C. R. (1971). Unified Theory of Linear Estimation, Sankhyā, Series A, Vol. 33, Part 4.

6. Rao, C. R. (1972). Some Recent Results in Linear Estimation, Sankhyā, Series B, Vol. 34, Part 4.

7. Rao, C. R. (1972). A Note on the IPM Method in the Unified Theory of Linear Estimation, Sankhyā, Series A, Vol. 34, Part 3.

8. Rao, C. R. (1973). Unified Theory of Least Squares. Communications in Statistics, Vol. 1, Number 1.

9. Rao, C. R. (1973). Representations of best linear unbiased estimators in the Gauss-Markoff model with a singular dispersion matrix. (J. Multivariate Analysis, in press).

10. Mitra, S. K. (1973). Unified least squares approach to linear estimation in a general Gauss-Markoff model (SIAM J. Appl. Math in press)

11. Mitra, S. K. and Rao, C. R. (1973). Projections under semi-norms and generalized inverse of matrices. (Tech. Report, Indiana University, Bloomington.

12. Rao, C. R. and Mitra, S. K. (1971). Further contributions to the theory of generalized inverse of matrices and its applications. Sankhya A, 33, 289-300.

13. Goldman, A. J. and Zelen, M. (1964). Weak generalized inverses and minimum variance linear unbiased estimation. J. of Research. Nat. Bureau of Standards, 68B, 151-172.